

## 4. Converse

---

*An essay by Arthur Sale, Emeritus Professor of Computer Science, University of Tasmania  
Fourth in a series of four, July 2011*

### Introduction

In this last of four essays, I am adopting the style of Galileo. Three friends are strolling along the bank of a river en route to a restaurant, discussing the issue of open access to scholarly research. SIMPLICIO<sup>1</sup> has been resurrected from the 1600s as the fall guy, asking the obvious questions. SALVOR<sup>2</sup> is the hard-headed doubter not certain what the goal is or how to achieve it. ARTOS is trying to address the problems. Fortunately, unlike Galileo, I do not need an *Imprimatur* nor a *Nihil Obstat* for this work.

### But it's done, right?

SALVOR: Well here we are with a problem. Open access seems so obvious and desirable, and will benefit everyone, but after a decade it hasn't happened!

SIMPLICIO: That isn't true is it? I read that great strides had been made and that the goal was finally in sight. Gold journals had grown by a large percentage! There are over 6,500 of them in DOAJ<sup>3</sup>.

ARTOS: Gold journals have indeed grown, and the data on them is pretty robust. Setting up a Gold journal is a pretty major undertaking, as is changing the business model of an existing journal. But Gold journals are by no means in the majority yet, or anywhere near it, nor are they the research outlets with the most prestige.

SALVOR: I guess the data on Gold journals is accurate unless a journal goes quietly bust without anyone noticing. There are also the fake Gold journals which exist to provide vanity publishing outlets for authors desperate to get a worthless paper published. There are a lot of examples of these in the third world, set up simply to extract money from gullible authors and institutions in author-side fees. I am sure some are in DOAJ.

SIMPLICIO: Of course we should not count these fake journals, which have arisen simply because a publishing fee is up-front. But I wasn't alluding to Gold journals, which I agree are not yet a critical point argument. What I meant was that there are so many mandates now, that we are close to having open access accepted as the norm. There are 132 such mandates registered in ROARMAP<sup>4</sup> and this is surely an under-estimate as not everyone registers their mandate.

---

<sup>1</sup> *Dialogue concerning the two chief World Systems*. Galileo Galilei. 1632.

<sup>2</sup> *Foundation*. Isaac Asimov. Mayor of the Foundation who did nothing, and resolved a crisis thereby.

<sup>3</sup> Directory of Open Access Journals, <http://www.doaj.org/>.

<sup>4</sup> Register of Open Access Repository Material Access Policies, <http://roarmap.eprints.org/>.

SALVOR: You are mistaken, SIMPLICIO. The data in ROARMAP is poor and your comment about registration is valid. But you fail to take into account three critical issues: policies, compliance and backsliding. Firstly, many of the so-called mandate policies have holes in them large enough to drive a bus through and this affects their performance. Secondly, even the best mandate does not achieve 100% compliance; generally a well-maintained policy might result in the deposit of 60% to 80% of the literature. Thirdly, and most significantly, ROARMAP collects 'intentions' and very often after a while the intention falls by the wayside, and the so-called 'mandate' is no longer in effect. If there is only weak pressure to add mandate records, there is certainly no pressure to remove mandate records.

ARTOS: Yes, you can see this in Australia. Quite probably the only general deposit mandate that has stood the test of time is that of Queensland University of Technology and that is largely because the person who instituted the policy is still in charge. At the University of Tasmania we have had three Deputy Vice-Chancellors of Research in that time, and their policies have flip-flopped all over the place, and not one of them has come close to a university wide mandate. I doubt there is any consensus on OA in Australia – the funding agencies are fence-sitters and the research institutions are more interested in their research evaluations by the Australian government.

SIMPLICIO: Well, I stand by my arguments. Mandates are working and universal open access through the Green Road is not far off.

ARTOS: May I propose some tests that could independently test your assertion? One is simple: identify a benchmark bunch of high quality journals, classify them by disciplines and check that they cover the fields we want to check, choose a time period, and then use the Internet to see which of the articles in the journals are accessible online, free, whether as an Accepted-Manuscript or a Version-of-Record. In other words, how much of the identified literature is indeed open access? You can use Google, Google Scholar, BASE, or any other tool you think fit. Repeat the exercise every year with the same group of journals.

SIMPLICIO: Well that seems fair. One could see if the open accessibility varies with time, or by discipline. But you had another test?

ARTOS: This is subject to more assumptions. We can know approximately how many articles one could expect from a university of a given size. Scopus data allows one to find country averages, and Australia collects audited lists of research outputs by institution and publishes them. This will give us an idea of how research outputs vary with size or prestige. Then apply these extrapolations of research outputs to all the institutions with mandates and assign them a score in 0-100% based on what was accessible, say for calendar years.

SALVOR: You mentioned assumptions?

ARTOS: Yes. One of them deals with aggregation of this information. I assumed that each deposit is a unique article. This may not be the case. An author may deposit several versions. If there are multiple authors, I assume that only one of them makes a deposit. This is possible if all the co-authors are in the same university, but unlikely if they are not. One can devise a protocol to minimize these effects. However, it does give us a view of what open access has achieved. Multiple authors should increase the probability of an article being open access.

## The Titanium Road

SIMPLICIO: I see where you are coming from. You think that open access is pretty well stagnating. Or at least not getting near a critical mass.

ARTOS: That's right. We need something more, and I believe that the something more requires addressing the set of players in this business that we have given up on. It is the *publishers* that are in control of the Gold Road. Repositories started off with researchers but this was rapidly found not to work, so attention turned to the *research managers*, the Green Road, and mandates. However, the people with the most to gain are still the *researchers* and *authors*. We should take a new look at their needs, and the Titanium Road provides such a review, in a way that has only just become possible.

SALVOR (pensively): But we tried to get the authors to deposit and they wouldn't in enough quantities. What's different now?

ARTOS: We made several mistakes, and that's one factor. The other is that social networking is now possible and much used – ordinary people use it as a way of promoting themselves and their views and think nothing of the time spent blogging and tweeting. How much more useful would be specialized scholarly social networking? And might there not be a large take-up in countries with good mobile or Internet connections? Best of all the Titanium Road can easily co-exist with the Gold Road and the Green Road – we don't need to abandon either. They are all routes to open access to articles.

SALVOR: I can see one of the mistakes. Asking researchers to deposit in an institutional repository looks like 'work', and the researcher would be unlikely to respond well. Even a small amount of extra effort would be resented, and of course that is all that deposit is. That explains the Arxiv situation, because it is not only a replacement for postal preprints, but it benefits the research, first, last and always. It is not driven by employers. So, ok, researchers won't do things for their employers readily, but were there other mistakes?

ARTOS: This is a difficult one. There were, and I refer to asking the researcher to deposit the final accepted draft (the Accepted-Manuscript). The natural tendency of the author is to deposit the Version-of-Record – it is what they published and what they stand by. We can only see this with hindsight. They consider themselves as the owner of the VoR, as well as the author. Technology has not completely erased the difference, although computer science conferences may be the exception in using camera-ready copy, and copyright rears its ugly head with the law as its usual ass. Asking the author to deposit an unnatural copy that they want to forget was and is a recipe for non-compliance. Boxes under my house are filled with copies of the paper Version-of-Record of the articles I have written, and I bet the same is true of most researchers. But I'd give them all away for an electronic archive, and indeed I have been working on that with a DVD and Mendeley. And that itself is a pale imitation of an archive of a CVBook in the cloud.

SALVOR: I want to come back to this copyright thing, but let's stay on track for a moment. Do you really think that people will adopt a Facebook-style archive in large numbers?

SIMPLICIO: Yeh! That bit is pretty obvious. All the young researchers will, and they have nothing much to put in their record to start it off. I can just see it, especially if you put Skype sessions and other semi-private information in the entry as an adjunct. And judging by Facebook, a

considerable number of oldies will take it up as well. There is nothing like polishing your personal presentation, is there? As long as it is easy, and that means a lot of automation. Another factor is the inbuilt Facebook trend of inviting ‘friends’ – in this new app we could have ‘collaborators’ and ‘cognates’, even ‘peers’ – well you come up with snappy words. I’m convinced about that aspect. A good scholarly ‘CVBook’ would have very many take-ups.

SALVOR: I still see a problem. That will mean that each co-author would have a copy of the article on their ‘book’. That would be a waste of space, and confusing. I’ve seen author lists as long as my arm!

ARTOS: Who cares about the number of copies on the Internet, where they reside, or how much space they take? Not I! The question of disk sectors or space is quite irrelevant. Indeed multiple authors implies multiple co-responsibilities, and a better guarantee of achieving open access through at least one of them! Yes, I have seen lists of co-authors up to 50 persons (in astronomy) and 20 (in biology) but they are no problem. Software can readily cope with multiple copies and Google does already. This is even easier if everyone uploads the Version-of-Record to their CVBook, because that is canonic and easily identified as identical.

## Copyright

SALVOR: I hate to bring this up, but I said I would, and you are implying that authors upload a copy of an article to their open access CVBook for which they have no rights. They signed away their copyright when they published their article. You can’t propose uploading a VoR.

ARTOS: I don’t have to advocate it or even say anything about it. Authors know little or nothing about copyright, but *before* their article is published they fear to do anything that might jeopardize their chances of being published. That’s why they sign the ridiculous publisher agreements. *After* publication they know the fact of publication is irrevocable and they have absolutely no compunction about giving away the article, even the Version-of-Record, or a photocopy, or a scanned file or a screen-dump. At all times they regard the article as theirs (and their co-authors jointly), and feel free to dispose of it *gratis* and *libré* as they feel fit.

SIMPLICIO: Here I agree with ARTOS. *My* article is *my* article, nobody else’s, certainly not a publisher’s. Also, I belong to several scientific mailing lists, and I have noticed that at least once a week if not more some member of the list posts a request for a toll access paper that they haven’t got. As far as I can tell, they nearly always get sent a copy of the publisher’s file, and not by the author! In many cases there is a ‘me-too’ factor as others clamour to be sent a copy as well. That’s thousands of papers bypassing toll-access, not a whisper from any publisher, and a totally obsolete set of laws. The whole scholarly publishing industry is in disarray legally, apart from its peer review role, and even that is provided to them free by us.

SALVOR: Not changing the subject, but had you noticed that besides authors not caring about copyright, neither do readers. Access is assumed, as is complaisant compliance by publishers. Indeed the only people that care about copyright are the publishers, who treat it with very little respect and never enforce it, and the research institutions to which it is a sacred cow that needs to be sacrificed to? This is probably a consequence of senior office as a research manager and badgering by people who know about the black-letter law of copyright.

ARTOS: Yes.

SIMPLICIO: Absolutely correct. The law is totally out of kilter with the practice here.

## **Integration**

ARTOS: Let's discuss something at a tangent, and that's discovery. If open access is to be a reality, it is not enough for every article to be accessible on the Internet, free. We also need it to be discoverable, by anyone, anytime, anywhere. Are there any issues here?

SALVOR: Well there is primarily the html harvesting by Google, Google Scholar, and other search engines, adding via software the pdf → html conversion. Then there are the OAI-PMH harvesters, and BASE is the main example. That would seem to capture everything possible.

SIMPLICIO: Yes, but it is not discriminatory. As a researcher I am not interested in seeing the metadata of an article and even its abstract is of minor value. I want to know that I have the article itself (free), in its entirety, and the Version-of-Record. I want to read it to see if it is relevant to my research. No existing search engine does what I want. The metadata is simply a help to discovery, of little use to me. I want a scholarly search engine which does something along the lines of Google Scholar, but is more effective.

SALVOR: Well if we had a CVBook system, we could expect it to harvest everything entered into CVBook and everything in DOAJ, and the institutional repositories and random websites would then be just a bonus to add. The discovery agent could also assign a reliability score, depending on the authority of the source and the number of copies. It would also give a 'PageRank' type of score based on the citations of the paper and the 'reliability' of the sources themselves – indeed the whole plethora of bibliometric measures could be used, built-in to the discovery engine. That's what I would want as a researcher. In fact I find it hard to envisage what might be possible when it does not yet exist.

ARTOS: It seems that we are unanimous that a social networking type of site for every academic author in the world would be a good thing, and that discovery could be implemented without too much trouble. Even competitive discovery engines could exist. It occurs to me that there is a conflict between the value of an article seen by the prestige of the authors or authors, by the JIF of the journal it appears in, by its PageRank, by its CitationRank, and who knows what other criteria. This would be invaluable in competitive assessment of reliability. The issue for me is how could this change take root amongst researchers. Where is the software?

## **Can it happen?**

SIMPLICIO: Surely the publishers have an interest in this? Perhaps they will create a social networking site for researchers and promote their own wares as well. It seems to me like an excellent route to an monopoly for them, and therefore efficiency, and therefore lower library subscription costs. The researchers need only post a link to their open access VoR on the publisher site rather than having to upload the whole document. Surely if the idea has merit, they'll claim the space for themselves?

SALVOR: SIMPLICIO, you amaze me. Publishers have no intention of making the VoR, which they regard as theirs, however trivial the changes they made, open access for free. If publishers did this,

they would simply channel your link into a per-copy fee site. And then what would we have achieved? Fortunately, so far they have not realized that this is an obvious possibility, being obsessed with their current business model. No, what we need is a new software, which we have begun to call *CVBook*, which has social networking of researchers at its core.

Quite what form *CVBook* should take needs a careful focus group to workshop out what researchers value and need, and what would motivate them to participate. I don't think I can guess what the outcomes will be by myself, but I am sure that it is possible.

ARTOS: So we need some new software funded by someone, and preferentially FOSS<sup>5</sup> under a GNU licence. I suppose JISC is a possibility though their expertise is more in dealing with institutions rather than individuals. But we could confidently expect that the funding bodies would get behind this. They have never liked mandates, but social networking between researchers with voluntary participation would be a really positive drawcard. And support from funding bodies is likely to have the force of 'law' for most researchers. To be fair, something similar is true for funder mandates, though not as strongly. The overt statement is not as persuasive as the supposed conspiracy bias in a voluntary suggestion that might affect the black art of the grant!

It will be difficult to guess what a scholarly Facebook might look like before starting work on it. Certainly, the place of residence is irrelevant, except insofar as competing versions must be sync-ed.

SALVOR (thoughtfully): so a personal record could live on a personal computer or iPad, or perhaps even an iPhone. It could be generated from an institutional repository, as long as the researchers keep interacting with it. But the cloud is the canonic version isn't it?

ARTOS: Yes. And any discussion about location is pointless. Many people think of the Internet as a central resource, and yet the *cognoscenti* know otherwise. The Internet is a distributed object with hundreds or thousands of name servers, even more interconnected routers, zillions of leaf computers, and the barest minimum of central control in the assignment of IP addresses and the protocols.

So conceptually what I conceive is a 'central' resource being the collection of *CVBook* entries. Each has whatever will be of value to the researcher, and that is where much more work is needed. Obviously, it must be possible to derive a traditional-type *curriculum vitae*, but the concept goes beyond that. Facebook 'friends' translate into a loose collaborative research group (say the tissue culture of pawpaw), and in widening circles with reducing personal choice and greater metadata use we move to broad areas of research (tissue culture of plants) into traditional disciplines (plant science, or tissue and cell culture generally). The linkages are a virtual map of what science there is, and is happening now. Searches can roam through these spaces, creating and using new temporary linkages.

Clearly editing a local copy on a personal PC does not require the Internet to be connected and that is an advantage, but as soon as connection occurs, the copy should be reconciled with the canonic cloud version, and until that is complete both versions will be marked as 'transitional'.

---

<sup>5</sup> **Free Open Source Software**, like Firefox, Open Office, GIMP, Linux and ePrints.

I put aside the security considerations, but they will be a key part of the process, hidden from and unknown to most researchers. Not only do we have to deal with spam, but also identity theft, and authentication.

SALVOR: So if this does not happen, then open access is destined for many more years of stagnation?

ARTOS: I fear so.

SIMPLICIO: I am convinced! Where can I sign up to be the first to create a CVPage? I'll promise to put the effort in to make my entry up to date! I really need to connect with all the important people in my field and I don't even think I know them all. My next paper needs expert comment on it before I submit it for publication, because I am sure that I have a proof for the Goldbach conjecture.

SALVOR: Oh SIMPLICIO, it just isn't that easy. But look, we have reached our destination and here are our companions. Are you not as famished as I? Let us not bore them over dinner.

